J|A|C|S
COMMUNICATIONS

Published on Web 08/29/2008

Native Conformation at Specific Residues in Recombinant Inclusion Body Protein in Whole Cells Determined with Solid-State NMR Spectroscopy

Jaime Curtis-Fisk, Ryan M. Spencer, and David P. Weliky*

Department of Chemistry, Michigan State University, East Lansing, Michigan 48824

Received May 26, 2008; E-mail: weliky@chemistry.msu.edu

In recent years there have been many advances in the production of recombinant proteins in bacterial hosts. A common problem is the sequestration of recombinant protein by the bacterium in an inclusion body which is the general term for a large insoluble protein aggregate. Harsh denaturing conditions are typically needed to solubilize the inclusion body protein with a consequent requirement of refolding to obtain functional protein. Despite these difficulties, inclusion bodies have some attractive features for recombinant protein production including each cell having one inclusion body which contains high purity protein and as much as 50% of the total mass of cellular protein.[1] Inclusion bodies are noncrystalline solids and there has been little high-resolution, residue-specific study to ascertain their molecular structure. The most common structural method has been infrared (IR) spectroscopy which provided information about the overall fractions of different types of regular secondary structure. These studies were typically carried out on dehydrated samples and suggested that there was more β sheet structure in the inclusion body protein than in the native fold. It was therefore proposed that the non-native β strands from different proteins associated together to form intermolecular β-sheets in a structure similar to amyloid protein.[2−8] Some reports indicate that a fraction of inclusion body protein is natively folded and functional.[3,9]

This paper describes application of solid-state NMR to determine residue-specific secondary structure in inclusion body protein. The methods are straightforward and inexpensive and should be broadly applicable to a wide variety of proteins in inclusion bodies. The methods were initially developed to study residue-specific conformation in membrane-reconstituted proteins and were based on the observation that in sequences of proteins of moderate size, a large fraction of the residues are in "unique sequential pairs", for example, there would only be one instance of a Leu followed by a Leu. In addition, NMR methods such as "rotational-echo double-resonance" (REDOR) can selectively detect the signal of the ^{13}C carbonyl (^{13}CO) nuclei which are directly bonded to ^{15}N nuclei, and there are well-known correlations between the backbone ^{13}CO NMR chemical shift of a residue and its local conformation.[10−12] Recombinant protein was therefore produced in *E. coli* cells in minimal media containing the 1-^{13}C amino acid and ^{15}N-amino acid of the respective N- and C-terminal residues of the unique sequential pair, and the filtered ^{13}CO NMR signal of the N-terminal residue of the pair was used to determine its conformation.[12]

These methods were first applied to "FHA2" which is a N-terminal 185-residue functional domain of the 221-residue "HA2" subunit of the influenza virus hemagglutinin protein.[12] FHA2 contains a ∼20-residue N-terminal "fusion peptide" that binds to the host cell membrane and plays a key role in infection. FHA2 is a membrane protein because of the fusion peptide while shorter constructs which lack the fusion peptide are soluble in aqueous solution and have been crystallized.[13] There is also a NMR structure of the HA2 fusion peptide in detergent micelles.[14]

One essential feature of the method was expression in minimal media containing labeled amino acids. However, growth only in minimal media provided <0.5 mg purified FHA2 per L culture and an alternate protocol was developed to first grow to high cell densities in rich media and to then switch to minimal media prior to expression.[12,15,16] Using this method, the purified yield of FHA2 from the soluble cell lysate was ∼10 mg/L culture. This was likely the fraction of FHA2 which was incorporated in cell membranes because the lysis buffer contained *N*-lauroylsarcosine detergent which is effective at solubilizing membrane proteins.

There was a significant amount of FHA2 in inclusion bodies as evidenced by a dominant FHA2 band in the SDS-PAGE gels of the unlysed fully hydrated *E. coli* cells and in the hydrated pellet formed from centrifugation of the insoluble material in the cell lysate, Figure 1a. In addition, ∼10 mg/L more FHA2 was purified after sonication of the aforementioned pellet in 8 M urea, and this denatured FHA2 could be refolded in detergent and reconstituted in membranes.[16] The relative band intensities in Figure 1a suggest that FHA2 is at least 10−20% of the total protein mass in the insoluble fraction and in unlysed whole cells and in the present study, the labeling/solid-state NMR method was applied to FHA2 in these two hydrated materials.[17,18] The ^{13}CO NMR signals were dominated by FHA2 as evidenced by comparison of the "S_0" (black) and "S_1" (red) REDOR NMR spectra of the pellet for which the labeling targeted the Leu-98/Leu-99 unique sequential pair, Figure 1b. The S_0 spectrum was composed of all the ^{13}CO signal in the sample whereas the S_1 spectrum lacked signal from ^{13}CO nuclei directly bonded to ^{15}N nuclei. The experimental S_1/S_0 integrated intensity ratio was 0.89 and correlated with the 0.92 ratio expected for fully labeled FHA2 with 13 Leus in the sequence. The experimental and (expected) ratios for labeling which targeted Gly-4 and Ala-7 were 0.93 (0.94) and 0.93 (0.91), respectively and were generally consistent with labeled amino acids being in the media only during the expression period and with FHA2 being the primary protein produced during the expression period.

The $S_0 - S_1$ REDOR difference spectrum primarily represents the filtered ^{13}CO signal of the N-terminal residue of the unique sequential pair and such spectra are presented in Figure 1 for (c, f, i) Gly-4, (d, g, j) Ala-7, or (e, h, k) Leu-98 and for FHA2 (c−e) purified and reconstituted in membranes, (f−h) in inclusion bodies in the insoluble cell lysate, or (i−k) in unlysed whole cells. There is general similarity of the spectra and peak ^{13}CO chemical shifts of the membrane-reconstituted, inclusion body, and whole cell FHA2 protein, Figure 1c−k and Table 1. The similarity of the insoluble fraction and the whole cell spectra correlated with most of the FHA2 being in inclusion bodies. In the existing structures, a N-terminal helix extends to residue 10 in the fusion peptide and there is a helix that extends from residue 38 to residue 105. The peak ^{13}CO shifts for Gly-1, Gly-4, Ala-7, and Leu-98 correlate with these helical conformations as evidenced by better agreement with distributions of backbone ^{13}CO shifts in helical conformation (175.5

 10.1021/ja8039426 CCC: $40.75

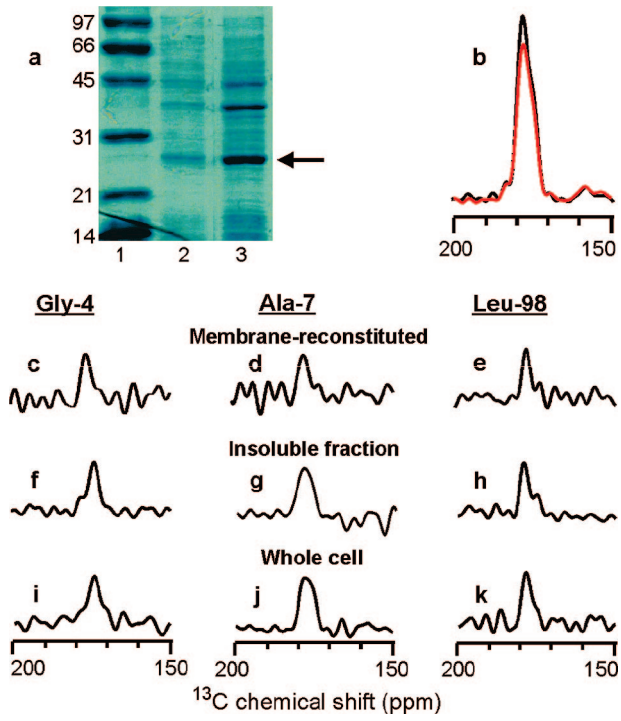

Figure 1. (a) SDS-PAGE gel: lane 1, molecular weight standards in kDa; lane 2, whole *E. coli* cells without lysis; and lane 3, the solid pellet obtained after centrifugation of the cell lysate. Different amounts of total protein were loaded in each lane and the arrow marks FHA2. (b−k) ^{13}C solid-state NMR spectra of FHA2. Panel b is for the insoluble fraction Leu-98 sample and the S_0 and S_1 REDOR spectra are displayed in black and red, respectively. The remaining panels are $S_0 - S_1$ difference spectra. Example expression media contained 1-^{13}C-Gly and ^{15}N-Ala (for Gly-4) or 1-^{13}C,^{15}N-Leu (for Leu-98). The most complete labeling of inclusion body FHA2 was obtained with addition of 30 mg/L of each labeled amino acid at the start of induction and at the 1 and 2 h times during the 3 h induction period. Experimental NMR conditions included 9.4 T spectrometer, 40 μL active sample volume, and sample cooling gas at −20 °C. Each membrane-reconstituted sample contained ∼5 mg FHA2 and each c−e spectrum was obtained in ∼3 days, while each of the other spectra was obtained in ∼1 $^1/_2$ days.

Table 1. Peak ^{13}C Carbonyl Chemical Shifts of FHA2 in ppm Units

sample type	residue			
	Gly-1	Gly-4	Ala-7	Leu-98
membrane-reconstituted	174.8	177.2	179.1	178.8
insoluble fraction	174.7	174.5	178.9	178.6
whole cell	175.2	174.7	177.9	178.4

± 1.2, 179.4 ± 1.3, and 178.5 ± 1.3 ppm for Gly, Ala, and Leu, respectively) than with characteristic shifts in β sheet/amyloid samples (170−172, 174−176, and 173−175 ppm, respectively).[5,6,11,19,20] The Ala-7 spectra of the insoluble fraction and whole cell samples are broader than the other spectra which may indicate some conformational heterogeneity at this residue. Overall, the data support a model in which native helical conformation is retained at least for a significant fraction of inclusion body FHA2. It is interesting that dominant β sheet chemical shifts were not observed in the fusion peptide region because this region is hydrophobic and could potentially form amyloid aggregates in the absence of membranes. Although ^{13}CO shifts are not absolutely definitive in determining conformation and are correlated with regions of the Ramachandran plot rather than precise dihedral angles, ^{13}CO shifts do provide significant conformational information particularly when helical shifts at several nearby residues are obtained (as in the fusion peptide region) or when sharp signals (indicating structural order) with very similar helical shifts are obtained in both the natively folded and inclusion body protein (as with Leu-98).[11]

The approach presented in this paper has two significant advantages over the more commonly used IR approach for study of inclusion body structure. First, conformation at individual residues is straightforwardly studied with the NMR method but only overall protein conformation is typically obtained with the IR method. It should therefore be possible to develop more detailed structural models with the NMR data. Second, the inclusion body and whole cell samples are fully hydrated with the NMR method rather than being dehydrated with the IR method. A large number of specifically labeled samples are required for a high-resolution NMR structural model of inclusion body protein but depending on linewidths, it may be possible to obtain structural data for samples with more significant labeling using methods previously applied to NMR structure determination of microcrystalline, amyloid, and membrane-associated proteins.[5,6,19−21] In summary, this study describes a general approach for residue-specific structural analysis of recombinant protein in inclusion bodies including those in whole *E. coli* cells as well as evidence for native conformation at some specific residues of a particular inclusion body protein.

Acknowledgment. The FHA2 plasmid was provided by Yeon-Kyun Shin and financial support was provided by Michigan State.

Supporting Information Available: FHA2 sequence and descriptions of cell growth and labeling, sample preparation, and NMR methods and analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Baneyx, F.; Mujacic, M. *Nat. Biotechnol.* **2004**, *22*, 1399–1408.
(2) Przybycien, T. M.; Dunn, J. P.; Valax, P.; Georgiou, G. *Protein Eng.* **1994**, *7*, 131–136.
(3) Oberg, K.; Chrunyk, B. A.; Wetzel, R.; Fink, A. L. *Biochemistry* **1994**, *33*, 2628–2634.
(4) Ami, D.; Natalello, A.; Gatti-Lafranconi, P.; Lotti, M.; Doglia, S. M. *FEBS Lett.* **2005**, *579*, 3433–3436.
(5) Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16742–16747.
(6) Ritter, C.; Maddelein, M. L.; Siemer, A. B.; Luhrs, T.; Ernst, M.; Meier, B. H.; Saupe, S. J.; Riek, R. *Nature* **2005**, *435*, 844–848.
(7) Carrio, M.; Gonzalez-Montalban, N.; Vera, A.; Villaverde, A.; Ventura, S. *J. Mol. Biol.* **2005**, *347*, 1025–1037.
(8) Umetsu, M.; Tsumoto, K.; Ashish, K.; Nitta, S.; Tanaka, Y.; Adschiri, T.; Kumagai, I. *FEBS Lett.* **2004**, *557*, 49–56.
(9) Garcia-Fruitos, E.; Aris, A.; Villaverde, A. *Appl. Environ. Microbiol.* **2007**, *73*, 289–294.
(10) Gullion, T.; Schaefer, J. *J. Magn. Reson.* **1989**, *81*, 196–200.
(11) Zhang, H. Y.; Neal, S.; Wishart, D. S. *J. Biomol. NMR* **2003**, *25*, 173–195.
(12) Curtis-Fisk, J.; Preston, C.; Zheng, Z. X.; Worden, R. M.; Weliky, D. P. *J. Am. Chem. Soc.* **2007**, *129*, 11320–11321.
(13) Chen, J.; Skehel, J. J.; Wiley, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8967–8972.
(14) Han, X.; Bushweller, J. H.; Cafiso, D. S.; Tamm, L. K. *Nat. Struct. Biol.* **2001**, *8*, 715–720.
(15) Murphy, O. J., 3rd; Kovacs, F. A.; Sicard, E. L.; Thompson, L. K. *Biochemistry* **2001**, *40*, 1358–1366.
(16) Curtis-Fisk, J.; Spencer, R. M.; Weliky, D. P. *Protein Expression Purif.* In press.
(17) Serber, Z.; Selenko, P.; Hansel, R.; Reckel, S.; Lohr, F.; Ferrell, J. E.; Wagner, G.; Dotsch, V. *Nat. Protoc.* **2006**, *1*, 2701–2709.
(18) Kim, S. J.; Cegelski, L.; Stueber, D.; Singh, M.; Dietrich, E.; Tanaka, K. S. E.; Parr, T. R.; Far, A. R.; Schaefer, J. *J. Mol. Biol.* **2008**, *377*, 281–293.
(19) Siemer, A. B.; Ritter, C.; Ernst, M.; Riek, R.; Meier, B. H. *Angew. Chem., Int. Ed.* **2005**, *44*, 2441–2444.
(20) Qiang, W.; Bodner, M. L.; Weliky, D. P. *J. Am. Chem. Soc.* **2008**, *130*, 5459–5471.
(21) Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420*, 98–102.

JA8039426

Supporting Information for "Native Conformation at Specific Residues in Recombinant Inclusion Body Protein in Whole Cells Determined with Solid-State Nuclear Magnetic Resonance Spectroscopy" by Jaime Curtis-Fisk, Ryan M. Spencer, and David P. Weliky

1. FHA2 sequence

GLF**G**AI**A**GFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
KLNRVIEKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAE**L**LV
ALENQHTIDLTDSEMNKLFEKTRRQLRENAEEMGNGSFKIYHKCDNACIE
SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDWVEHHHHHH

The first 185 residues in the sequence are the 185 N-terminal residues from the HA2 protein of the influenza virus, X31 strain, and the last 8 residues are non-native. The sequence is written in lines containing 50 residues and the Gly-1, Gly-4, Ala-7, and Leu-98 residues are in bold and underlined.

2. Expression and labeling of FHA2

The protocol for expression and labeling of FHA2 in inclusion bodies was similar to that described for FHA2 incorporated in bacterial membranes and subsequently purified in detergent and reconstituted in synthetic membranes, cf. Curtis-Fisk, J.; Preston, C.; Zheng, Z. X.; Worden, R. M.; Weliky, D. P. *J. Am. Chem. Soc.* **2007**, *129*, 11320-11321 and Curtis-Fisk, J.; Spencer, R. M.; Weliky, D. P. *Prot. Expr. Purif.* in press. Dr. Yeon-Kyun Shin at Iowa State University donated the FHA2 plasmid which contained the Lac promoter and kanamycin resistance and the plasmid was transformed into E. coli BL21(DE3) cells. The cells were first grown overnight to an OD_{600} of ~8 in 0.5 L media containing luria broth and 5 g glycerol. The cells were then switched into 0.5 L minimal media that contained 5 g glycerol. After resumption of growth in this media

for one hour, 30 mg aliquots of one or two labeled amino acids were added to the media and FHA2 expression was induced for three hours with 1 mM isopropyl thiogalactoside. The labeled amino acids were: (1) Gly-1 sample, 1-^{13}C Gly and ^{15}N-Leu; (2) Gly-4 sample, 1-^{13}C Gly and ^{15}N-Ala; (3) Ala-7 sample, 1-^{13}C Ala and ^{15}N-Gly; and (4) Leu-98 sample, 1-^{13}C, ^{15}N Leu. Expression was done at 37 ^{o}C rather than room temperature to augment production of inclusion bodies, cf. Vera, A.; Gonzalez-Montalban, N.; Aris, A.; Villaverde, A. *Biotechnology & Bioengineering* **2007**, *96*, 1101-1106. Because the bacterial membrane space is limited, a greater fraction of FHA2 will be incorporated into membranes at the beginning of the expression period and a greater fraction will be in inclusion bodies at the end of the expression period. Complete labeling of the inclusion body FHA2 therefore required additional aliquots of 30 mg/L of labeled amino acid at the one and two hour times during the expression period. After three hours of expression, the cells were centrifuged and the whole cell NMR sample was taken from the cell pellet. The cell yield with this protocol was typically 10 g cells/L culture.

3. Separation and native purification of the bacterial membrane fraction of FHA2 and reconstitution of this fraction in synthetic membranes

The protocol for separation and native purification of the fraction of FHA2 incorporated in bacterial membranes has been previously described, cf. Curtis-Fisk, J.; Preston, C.; Zheng, Z. X.; Worden, R. M.; Weliky, D. P. *J. Am. Chem. Soc.* **2007**, *129*, 11320-11321 and Curtis-Fisk, J.; Spencer, R. M.; Weliky, D. P. *Prot. Expr. Purif.* in press. The protocol relied on solubilization of this fraction in a lysis buffer that contained 0.5% w/v *N*-lauroylsarcosine detergent and subsequent affinity purification based on the non-native histidine tag. Increased lysis time did not increase the amount of FHA2

purified from the soluble cell lysate and *N*-lauroylsarcosine is a very good detergent for solubilizing proteins in the bacterial membranes. The lysis was therefore considered to be complete in terms of breaking the cells and cell membranes and solubilizing the FHA2 which was embedded in the cell membranes. Membrane reconstitution of purified FHA2 also followed the previously published protocol. Briefly, a 5 mL solution was prepared which contained ~5 mg FHA2, 160 mg n-octyl-β-D-thioglucopyranoside detergent, 32 mg di-*O*-tetradecylphosphatidylcholine (DTPC) lipid, and 8 mg di-*O*-tetradecylphosphatidylglycerol (DTPG) lipid. The detergent was then removed by dialysis against a pH 5.0 buffer solution. The resultant membrane bilayer suspension with bound FHA2 was centrifuged and the pellet was transferred to a 4 mm diameter magic angle spinning (MAS) rotor that had ~40 μL active sample volume.

4. Preparation of inclusion body samples

Unlike the fraction of FHA2 incorporated in the bacterial membranes, the inclusion body FHA2 was not solubilized by the lysis buffer and was in the pellet obtained from centrifugation of the cell lysate at 50000*g* for 30 minutes. While optimizing the inclusion body protocol, pellets were also obtained with smaller centrifugation *g* forces and shorter times, but the SDS-PAGE gels of these pellets were very similar to the gel displayed in lane 3 of Fig. 1 in the main text. In addition, NMR samples were taken from different regions of the same pellet but the NMR spectra of these different samples were very similar. For the final protocol, the insoluble fraction NMR sample was taken from all regions of the pellet.

5. Solid-state nuclear magnetic resonance (NMR) experiments and analysis

Solid-state NMR rotational-echo double-resonance (REDOR) experiments were carried out using a 9.4 T spectrometer and a 4.0 mm MAS probe tuned to ^{13}C detection at 100.8 MHz, ^{1}H decoupling at 400.8 MHz, and ^{15}N dephasing at 40.6 MHz. The pulse sequence was ^{1}H-^{13}C cross-polarization followed by a REDOR dephasing period and ^{13}C-detection with ^{1}H decoupling during the latter two periods. Data were alternately acquired without (S_0) and with (S_1) the ^{15}N π pulses during the dephasing period and respectively represented the full ^{13}C signal and the ^{13}C signal minus ^{13}Cs close to ^{15}N nuclei. The experimental MAS frequency was 8.0 kHz, the dephasing time was 2.0 ms, the recycle delay was 1.0 s, and the ^{13}C chemical shifts were externally referenced to the methylene peak of adamantane at 40.5 ppm which allows direct comparison to databases of ^{13}C chemical shifts as a function of conformation, cf. Zhang, H. Y.; Neal, S.; Wishart, D. S. *J. Biomol. NMR* **2003**, *25*, 173-195 and Morcombe, C. R.; Zilm, K. W. *J. Magn. Reson*. **2003**, *162*, 479-486. Other details of the NMR experiments have been previously described, cf. Curtis-Fisk, J.; Preston, C.; Zheng, Z. X.; Worden, R. M.; Weliky, D. P. *J. Am. Chem. Soc.* **2007**, *129*, 11320-11321. Peak ^{13}C carbonyl chemical shifts were measured with ±0.3 ppm precision. During data acquisition, samples were cooled by flowing nitrogen gas whose temperature was –20 ^{o}C. For the inclusion body and whole cell samples, the NMR probe could not be tuned when the cooling gas temperature was higher than –20 ^{o}C presumably because of effects from the ~200 mM total salt concentration in these samples. This reasoning is evidenced by: (1) no problems with tuning the membrane-reconstituted samples at temperatures above –20 ^{o}C; and (2) the ~10 mM salt concentration in the membrane-reconstituted samples. In the future, it will

be possible to do the inclusion body and whole cell NMR experiments at temperatures higher than −20 ^{o}C using recently developed NMR probes that minimize electric effects in conductive samples, cf. Stringer, J. A.; Bronnimann, C. E.; Mullen, C. G.; Zhou, D. H. H.; Stellfox, S. A.; Li, Y.; Williams, E. H.; Rienstra, C. M. *J. Magn. Reson*. **2005**, *173*, 40-48.

Because the FHA2 contained a ^{13}CO/^{15}N unique sequential pair, the $S_0 - S_1$ difference was predominantly the filtered signal of this pair, cf. Yang, J.; Parkanzky, P. D.; Bodner, M. L.; Duskin, C. G.; and Weliky, D. P. *J. Magn. Reson.* **2002,** *159,* 101-110. For spectra of labeled FHA2 which targeted the ^{13}CO of Gly-4, Ala-7, or Leu-98, the experimental S_1/S_0 integrated intensity ratio was respectively 0.93, 0.93, or 0.89 and correlated with the expected ratios of 0.94, 0.91, and 0.92 for putative 100% labeling. These expected ratios were based on an approximate model in which the S_1/S_0 intensity ratio was: (number of FHA2 residues of the same type as the ^{13}CO labeled amino acid minus one)/(number of FHA2 residues of the same type as the ^{13}CO labeled amino acid), e.g. 10/11 for Ala. The $S_0 - S_1$ signal for the sample targeting Leu-98 also had some contribution from the other Leus (~10% per Leu) in the sequence because of the 2.5 Å intra-residue ^{13}CO $^{\cdots}$ ^{15}N distance in 1-^{13}C, ^{15}N-Leu, cf. Yang, J. Ph. D. thesis, Michigan State University, East Lansing, MI, 2003. These contributions are probably responsible for the experimental S_1/S_0 integrated intensity ratio which is smaller than the ratio predicted by the approximate model. Although some of these intra-residue Leu contributions would likely have helical shifts similar to the shift of Leu-98 (presuming that FHA2 remains folded), the contributions from other Leu residues would be dispersed over a range of non-helical shifts and would be less apparent in the overall appearance of

the difference spectrum. This idea is supported by Fig. 2b in which there is a large apparent difference between the S_0 and S_1 intensities in the 177.5-179.5 ppm helical shift region and much smaller difference in the non-helical lower shift region.

There will also be natural abundance contributions to the $S_0 - S_1$ difference signal. A model is developed for these contributions in the context of the following approximations: (1) the recombinant protein makes up a large portion of the mass of the sample; and (2) only the recombinant protein is labeled and the labeling is close to 100%. These approximations appeared to hold for the insoluble fraction of the cell lysate with FHA2 labeling which targeted the Gly-4 or Ala-7 residues. The recombinant protein is considered to have "*D*" residues that are ^{13}CO labeled and "*E*" residues that are ^{15}N labeled. There is a single unique sequential ^{13}CO-^{15}N labeled pair and this pair contributes 1.0 intensity to the $S_0 - S_1$ difference signal. The approximate contribution to the difference signal intensity by natural abundance ^{15}N dephasing of the $D - 1$ other labeled ^{13}COs will be $(D - 1) \times (1.1) \times 0.0037$ where the 1.1 factor considers the effect at 2 ms dephasing time of ^{15}N separated by one and two bonds from the labeled ^{13}CO and 0.0037 is the ^{15}N fractional natural abundance. Similarly, the approximate contribution to the difference signal intensity from natural abundance ^{13}COs dephased by the $E - 1$ other labeled ^{15}Ns is $(E - 1) \times (1.1) \times 0.0111$. The fractional natural abundance contribution to the total $S_0 - S_1$ signal intensity is therefore:

$$\{[(D - 1) \times 0.00407] + [(E - 1) \times 0.0122]\}/ \\ \{1.0 + [(D - 1) \times 0.00407] + [(E - 1) \times 0.0122]\}$$

For FHA2 with labeling which targeted the Gly-4 residue, $D = 16$, $E = 11$, and the ratio was 0.15. A much larger natural abundance contribution to the $S_0 - S_1$ difference signal will be observed if: (1) the recombinant protein is much larger than FHA2 and *D* and *E*

are much larger; (2) the recombinant protein is only a small fraction of the sample; and/or (3) proteins other than the recombinant protein are labeled during the expression period and are insoluble in the cell lysate.

In the context of the model, the lineshape of the natural abundance contribution to the $S_0 - S_1$ difference signal will primarily depend on the amino acid identities and the conformations of the residues in sequential pairs containing labeled residues. Because there will likely be a wide variety of amino acid types and perhaps conformations for these residues, there will likely be significant chemical shift dispersion in the lineshape. For large values of D and E, the $S_0 - S_1$ difference signal would therefore appear to contain a sharp signal from the ^{13}CO in the unique sequential pair and a broad signal from the contributions described in the aforementioned model.

Each of the membrane-reconstituted samples contained ~5 mg FHA2 and the corresponding spectrum was obtained with ~3 days of signal averaging. The signal-to-noises of the insoluble fraction and whole cell samples were typically comparable or higher than that of the corresponding membrane-reconstituted sample and were obtained in ~1½ days, cf. Fig. 1 in the main manuscript. There may be variation in motion and hence cross-polarization and REDOR efficiencies for the different sample types which would differentially affect signal intensities, but neglecting this potential variation, the signal-to-noises in the spectra suggest that the insoluble fraction and whole cell NMR samples contained 5-10 mg FHA2. The NMR active sample volume was ~40 μL which could likely contain ~50 mg of material. The overall analysis therefore suggested that 10-20% of the mass in the insoluble fraction and whole cell NMR samples was FHA2 and

this percentage was generally consistent with the apparent mass fraction of FHA2 in the gels in Fig. 1a of the main manuscript.

In the main manuscript, the reported distributions of ^{13}CO chemical shifts for Gly, Ala, and Leu residues in helical conformations were obtained from Table 7 in Zhang, H. Y.; Neal, S.; Wishart, D. S. *J. Biomol. NMR* **2003**, *25*, 173-195. The distributions were based on ~180 proteins for which both the ^{13}CO assignments and high-resolution structures were known. Inclusion of a residue in the helical distribution was based its hydrogen bonding and on its values of the backbone dihedral angles φ and ψ with typical helical ranges being $-120^{o} < \varphi < -34^{o}$ and $-80^{o} < \psi < -6^{o}$. These are generally consistent with the experimental ranges of dihedral angles in α helices, cf. Hovmoller, S; Zhou, T; Ohlson, T. *Acta Cryst*. **2002**, *D58*, 768-776. In the main manuscript, the reported ^{13}CO chemical shift ranges for β sheet/amyloid conformation were based on solid-state NMR-derived ^{13}CO shifts of the membrane-associated HIV fusion peptide, β amyloid peptide, and HET-s prion protein reported in references 5, 6, 19, and 20 from the main text. An increment of 2.0 ppm was added to the shifts from references 5 and 19 so that they would be referenced in the same way as the shifts in the helical distributions. The β sheet ^{13}CO shifts from these samples in ppm units were: Gly, 171.5, 170.3, 170.7, 171.4, 170.7, 172.0, 172.0; Ala, 174.2, 174.9, 175.6, 176.9, 174.7, 176.0, 175.2, 173.3; and Leu, 174.2, 174.7, 174.4, 174.8, 175.0, 173.0.